Appendix B

Alternative Model for Diesel Cancer Risk Assessment

B.1. INTRODUCTION

As previously discussed in Chapter 11, the most appropriate method to assess cancer risk of diesel exhaust is to take into account effects of both particles (carbon core) and organics because evidence exists that both agents are involved in carcinogenic process. The reasons for this conclusion are based on the following observations: (1) organics include a variety of polycyclic aromatic hydrocarbons (PAHs) and nitroaromatics, many of which are known to be carcinogenic; (2) the results of recent studies on inert particles and carbon black in rats strongly support the hypothesis that the carbon core of the diesel particle may be the primary component responsible for the induction of lung cancer; (3) PAHs are unlikely responsible for all observed tumors because they account for less than 0.1 μg/mg particulate matter (Tong and Karasek, 1984); and (4) the observation of disproportionate high tumor incidence in high exposure animals coincides with disproportionate increase of cumulative lung burden of diesel particle as exposure concentration increases.

A workshop on Research Needs for Risk Assessment of Inhaled Particulate Matter was organized and sponsored by the U.S. Environmental Protection Agency (EPA) in March, 1992. The purpose of the workshop was to determine the extent of information that can be used for quantitative risk assessment and to discuss mechanisms of particle-induced lung tumors to serve as a guidance for future research needs. Two major, among several other, conclusions that are relevant to quantitative risk assessment were reached by the Workshop:

(1) particle overloading of the lung tissue may induce both initiation (by PAH specific adducts and adducts through oxygen radicals) and cell proliferation steps in tumor formation, and

(2) more research is needed to improve the risk assessment of particle-induced lung cancers.

Although there are not enough data available to construct a biologically based dose-response model, it is desirable to investigate implications of the hypothetical mechanisms proposed by the workshop. The purpose of the alternative modeling presented in this report is to do just that. Briefly, the biological issues and their implications to quantitative risk assessment that we would like to consider are the following.

1. Particles deposited in lung are phagocytized by alveolar macrophages. Because the phagocytizing macrophages in animals from high-dose group may be more likely to be activated to release mediators including reactive oxygen species, cytokines, and growth factors, it is of interest to determine whether or not the available tumor response data are

consistent with the hypothesis that the particle burden affects both initiation and proliferation in carcinogenic process.

2. Organic materials can also induce specific adducts which may contribute to cell initiation. However, given its low content, the contribution of organics to tumor induction may be very small. Can a dose-response model that is consistent to the proposed biological concept be constructed with both organics and particles as dose metrics?

3. If a model that has the above biological interpretation and is consistent with the bioassay data can be constructed, what would be its implications on quantitative risk assessment of diesel exhaust emissions, and how would its results compare with those predicted by the linearized multistage (LMS) model?

B.2. PRELIMINARY CONSIDERATIONS

In order to evaluate the impact of various biological assumptions on diesel risk assessment, it is necessary to construct a mathematical dose-response model that takes into account the biological mechanisms discussed in the EPA workshop. Because an issue of significant importance in diesel risk assessment is the effect of lung overloading on tumor induction, the model should possess the fol/lowing properties.

1. It should depend on both types of dose metrics: organics, and carbon core. It should allow one to account for the contribution of organics and carbon core individually and/or jointly to tumor induction/formation.

2. It should allow for the possibility that model parameters can change with time because of the increasing lung burden during exposure.

3. The cell proliferation and tumor induction/formation should be stochastic in nature; it is not realistic to assume deterministic clonal growth. For instance, it should not be required to assume that all cells divide at the same age.

To accomplish these goals, we assume that a normal cell can be initiated by both organics and carbon core. Denote the initiation rate by μ_1, which is a function of background and diesel-induced rates (as specified below). Because an initiated cell (I-cell) eventually either goes into cell death, or enters the cell cycle (including cells in quiescence, G_0), it is reasonable to assume that the cell lifetime for an I-cell follows certain probability distribution. Under this model, a cell in G_0 phase is equivalent to the case where it has a very long lifetime with certain probability (i.e., in the right-hand tail of the cell lifetime distribution). At the end of an I-cell's lifetime, it either dies (death) with probability β, divides into two daughter cells (birth) with probability α, or divides into one I-cell and one malignant cell (second transition) with probability μ_2; $\alpha + \beta + \mu_2 = 1$. Instead of assuming that a single malignant cell is equivalent to a tumor as in the MVK model

proposed by Moolgavarkar and colleagues (1979, 1981), we assume that a malignant cell has a certain probability to become a tumor; this probability is assumed dose-dependent, thus allowing for an evaluation of dose effect on tumor progression. It should be noted that the proposed model does not exclude the possibility that it may take more than one step (for a normal cell) to become "initiated." The rate of initiation used in the model should be viewed as a net rate which represents several genetic alterations and repairs.

B.3. MATHEMATICAL MODEL AND PARAMETERS ESTIMATION

We shall proceed to construct a dose-response function P(t:d,D), probability of cancer by time (age) t, which depends on both organic, d, and particle (carbon core), D, and incorporates the biological concept discussed previously. Because the model parameters that are not directly observed in laboratory can only be statistically estimated from high concentration cancer bioassay data, the model constructed should not be considered a valid model of diesel-induced carcinogenesis; uncertainty about the low-dose extrapolation still remains. Some discussions about the need for further laboratory measurements will be given later.

The model with the desirable features discussed above falls into one of several classes of stochastic models that have been developed by EPA's Office of Health and Environmental Assessment (OHEA): namely, the stochastic model which was originally proposed by Chen and Farland (1991) and extended into one with time varying parameters by Tan and Chen (1992). This model will be used as the basis for constructing a biologically based dose-response model. A brief mathematical description is presented in Appendix B-2.

The time to event data from Mauderly et al. (1987) are used to estimate model parameters. The data from Mauderly et al. are useful because they contain information on natural mortality and serial sacrifice of animals with or without (malignant) tumors, valuable information for estimating tumor latency. To utilize the information from serial sacrifice in Mauderly et al. an (E-M) algorithm is derived (see Appendix B-1) and used to calculate maximum likelihood estimates of parameters.

B.3.1. Model Parameters and Notations

The following parameters are incorporated in the dose-response model, which includes initiation rate (μ_1), proliferation rate ($\gamma\alpha$), conversion rate ($\gamma\mu_2$), and probability of progression (q). The death rate for the initiated cells is implicitly defined by $\gamma(1 - \mu_2 - \alpha)$. These parameters are all dose dependent.

D: Dose of carbon core, mg/cm^2 of lung epithelial surface; D varies over time

d: Dose of organics, mg/cm^2 of lung epithelial surface

μ_1: Dose-related initiation rate (per cell per day) that depends on μ_0 (background rate), d, and D by $\mu_1 = \mu_0(1 + ad + bD)$; a and b are paramaters to be estimated.

μ_2: Probability of producing a malignant cell at the end of an initiated cell (I-cell) lifetime

α: The probability that an I-cell divides into two daughter cells at the end of its lifetime

q: Probability that a single malignant cell will develop into a malignant tumor

γ: $1/\gamma$ is the mean I-cell lifetime in days; a cell lifetime ends if it either goes into mitosis, or cell death. Note that if one assumes that the probability for a cell to get into mitosis is about the same as cell death then the mean cell lifetime can be conveniently interpreted as time to mitosis (i.e., cell turnover time); thus, shorter cell lifetime implies more frequent cell division. Note that the time to mitosis is a random variable here, not a fixed constant as in the assumption made in the Greenfield et al. (1984) model that has been used extensively by Cohen and Ellwein (1988) to analyze experimental bladder cancer.

N: Number of (normal) target cells

B.3.2. Practical Considerations

By statistical theory alone the E-M algorithm developed in this report provides an elegant procedure which can be used to test hypotheses whether a particular parameter is influenced by organics and carbon core individually or both together. For instance, one could postulate that the parameter γ (reciprocal of which represents mean cell lifetime) is given by $\gamma(d,D_i) = \gamma_0 + \gamma_{11}d + \gamma_{12}D_i$, and then proceed to test a null hypothesis that $\gamma_{11} = 0$, no effect of organics on cell lifetime. This temptation, however, must be resisted because there would be too many parameters that must be estimated if such statistical tests are to be performed. Therefore, rather than performing such a statistical exercise, we proceed with a biologically plausible assumption that parameters q and γ depend only on lung burden of carbon core, C.

The duration of the Mauderly et al. study was about 940 days. To construct a dose-response model with time-dependent lung burden, the time interval (0,940] is divided into five subintervals; each subinterval spans 6 mo except for the last subinterval, which spans from 730 (2 years) to 940 days. Corresponding to an ambient air concentration of diesel emissions in mg/m^3, the deposition-retention model developed by Yu et al. is used to calculate dosimetric (d, D_i), I = 1, 2, ..., 5, where organics dose, d, is not changing with time because it reaches steady state quickly after exposure begins and D_i is the lung burden of carbon core during the ith subinterval.

The assumptions about dose-parameters relationship are given below.

1. The initiation rate associated with a lung burden $\{d, D_i, I = 1, 2, ..., 5\}$ is given by $\mu_1(d,D_i) = \mu_0(1 + a * d + b * D_i)$, for I = 1, 2, ..., 5. This is the only parameter that is assumed to depend on both d and D.

2. Probability of tumor formation from a malignant cell is assumed to be dependent on lung burden D by $q(D_i) = q_0 + q_1 D_i$, I = 1, 2, ..., 5. To simplify calculation, the possibility that q is also dependent on organics d is not considered.

3. The cell lifetime parameter γ is assumed related nonlinearly to lung burden D by $\gamma(D_i) = \gamma_0 + \gamma_1 \text{Log}(1 + D_i)$, I = 1, 2, ..., 5.

To reduce the number of parameters that must be estimated from the Mauderly data, some of the background parameters (μ_0, q_0, and γ_0) for the dose-response model are estimated from the National Toxicology Program (NTP) historical control rate on Fischer-344 rats (reconstructed from Portier et al., 1986). Giving these background parameters, the dose-related parameters are then estimated by the E-M algorithm, which is described in Appendix B-2. Using tumor response data from Mauderly et al. (1987) and the corresponding dosimetric in Table B-1, the resultant parameter estimates for the model are given in Table B-2. To have some appreciation about the implication of the Mauderly et al. (1987) study, the estimated initiation and proliferation (for I-cells) rates for the study are given in Table B-3. Although these values may not represent reality (because they are not actual laboratory measurements), they could be used as a guidance for future research planning. For instance, Table B-3 (along with a discussion about Table B-7) suggests that a slight increase of proliferation rate could cause a drastic increase on tumor incidence, but only if the initiation rate is high enough. This conclusion seems to suggest that although the promotion effect of growth factors is important for tumor induction, the initiation effect of carbon core and/or organics is also essential.

Table B-1. Dosimetric (mg/cm^2 lung surface) use in modeling[a]

Exposure group	d	D_1	D_2	D_3	D_4	D_5
0.35	2.5E-6	6.23E-5	8.75E-5	8.97E-5	9.02E-5	9.02E-5
3.5	3.6E-5	7.54E-4	2.40E-3	3.91E-3	5.25E-3	6.29E-3
7.08	7.3E-5	1.98E-3	5.49E-3	8.56E-3	1.12E-2	1.44E-2

[a]d is organics; D_i, I = 1, 2, ..., 5, are average lung burden of carbon core over five time intervals. These values are calculated by Yu et al. retention model in Appendix C.

Table B-2. Maximum likelihood estimates for model parameters

Parameter[a]	Estimate
μ_0	1.033E-7
a	1.103E+4
b	3.214E+2
μ_2	7.907E-7
q_0	1.035E-1
q_1	5.332E-2
γ_0	1.662E-2
γ_1	2.647
α	5.443E-1
N[b] (given)	8.80E+7

[a]Background parameters μ_0, q_0, and γ_0 are estimated separately from NTP historical control data.
[b]The number of target cells N is assumed to be 10-fold of Type II cells in mice, which is given in Kauffman (1974). It is not essential for N to be given accurately because $N\mu_0$ appears as a single term in the model; the estimated μ_0 will compensate for the under- or over-estimation of N.

Table B-3. Relative magnitude of initiation $\frac{\mu_1(d,D)}{\mu_0}$ and proliferation ($\Gamma[D]/\Gamma_0$) potential for the exposed versus control groups in Mauderly et al. (1987) study

		Exposed Groups		
	Time Interval	Low	Mid	High
Initiation	1	1.048	1.639	2.442
	2	1.056	2.168	3.570
	3	1.056	2.654	4.556
	4	1.056	3.084	5.405
	5	1.056	3.419	6.433
Proliferation	1	1.004	1.052	1.137
	2	1.006	1.166	1.379
	3	1.006	1.270	1.590
	4	1.006	1.362	1.770
	5	1.006	1.424	1.989

B.4. RESULTS

The dose-response model predicts only probability of tumor occurrence (i.e., tumor incidence) by time t due to an exposure scenario. Because the probability of tumor occurrence is not directly observable (note that when an animal dies with a tumor, it only tells us that a tumor occurred before that time), the model can not be evaluated against the observed data. (Although it is possible to use the dose-response model, with additional assumptions, to calculate tumor mortality, we prefer to evaluate the dose-response model alone because it will be used to predict cancer risk). To evaluate the reasonableness of the model constructed, it is possible only to compare the observed tumor mortality rate (after adjusting for the competing risk) and the predicted tumor incidence rate, with the understanding that observed values are expected to be smaller (this is particularly true at an early stage of tumor development when a tumor is small) than the predicted tumor incidence. Table B-4 appears to bear this out; all of the predicted tumor incidences are either greater than the observed tumor mortality rates, or within the confidence bounds calculated from the observed tumor mortality. The observed tumor mortality rate is calculated by life-table approach. The probability of tumor mortality can only be calculated up to about 900 days because after 900 days tumors are no longer discovered by natural mortality only; in fact, the majority of tumors are discovered by sacrifice.

B.5. RISK PREDICTIONS UNDER VARIOUS EXPOSURE SCENARIOS

For comparison, excess lifetime risks (see Tables B-5 and B-6) due to various exposure scenarios are calculated by the alternative model and the linearized multistage (LMS) model. Both point (maximum likelihood estimate) and 95% upper bound estimates are provided for the alternative model, whereas only upper bound estimate is provided for the LMS model because its linear component (which is notoriously unstable) is estimated to be 0. The 95% upper bound for the alternative model is calculated by the same approach as for the LMS model; (i.e., by increasing parameters a and b until the log-likelihood exceeds a critical value). To extrapolate from animal-based risk estimates to human, two assumptions are made: (1) lung burden in terms of $\mu g/cm^2$ of lung surface is equally potent between animals and humans, and (2) 6 mo of animal life is equivalent to 18 years of human life. The latter assumption is necessary because life-span must be divided into five subintervals to account for different parameter values.

Table B-5 compares predicted risks for humans due to continuous exposure (24 h/day) calculated by alternative and LMS models. It is interesting to see that risk calculations under various exposure concentrations are very similar using the two different models. Table B-6 gives excess risks due to exposure to 2.6 $\mu g/m^3$ of diesel emissions, 16 h/day, 7 days/week; and 15

Table B-4. Comparison of observed tumor mortality rate and predicted probability of cancer occurrence by time (T) when a (malignant) tumor was observed in rats

Exposure (mg/m^3)	Time tumor observed (days)	Observed tumor mortality rate[a] (95% C. I.)		Predicted tumor rate by time (t)
Control	538	0.0051	(0, 0.015)	0.002
	551	0.010	(0, 0.028)	0.003
0.35	710	0.007	(0, 0.022)	0.005
	863	0.025	(0, 0.063)	0.008
3.50	891	0.016	(0, 0.126)	0.039
	895	0.036	(0, 0.052)	0.040
7.08	646	0.006	(0, 0.019)	0.039
	672	0.013	(0, 0.037)	0.046
	701	0.021	(0, 0.041)	0.054
	729	0.027	(0, 0.059)	0.064
	742	0.039	(0.004, 0.075)	0.069
	798	0.052	(0.009, 0.095)	0.097
	810	0.066	(0.016, 0.115)	0.104
	839	0.081	(0.023, 0.138)	0.123
	840	0.096	(0.032, 0.161)	0.123
	847	0.112	(0.041, 0.183)	0.128
	856	0.129	(0.052, 0.207)	0.135
	859	0.146	(0.063, 0.229)	0.137
	883	0.168	(0.077, 0.259)	0.156
	895	0.191	(0.091, 0.291)	0.166

[a]Calculated by the life table procedure. Note that observed values are mortality, which are expected to be smaller than the (predicted tumor) incidence. This expectation is particularly true at early stage of tumor development when a tumor was small.

$\mu g/m^3$, 8 h/day, 5 days/week. The concentration 2.6 $\mu g/m^3$ was reported by EPA's Office of Mobile Sources to be the annual mean exposure of the U.S. population to diesel particulate matter in 1986 and is only slightly higher than the most recent estimate of 2.03 $\mu g/m^3$ in an EPA draft document (Motor Vehicle-Related Air Toxic Study, April, 1993); the concentration 15 $\mu g/m^3$ was reported to be the particulate exposure for workers on urban freeways in an EPA report by Carey (Air Toxics Emissions from Motor Vehicles, 1987, EPA-AA-TSS-PA-86-5).

Table B-5. Comparison of excess risk for humans due to continuous exposure of various concentrations of diesel exhaust emissions under two different models

	Alternative model[a]		
Exposure concentration ($\mu g/m^3$)	MLE	95% u.b.	LMS model[b]
0.01	7.68E-8	1.35E-7	1.71E-7
0.1	8.12E-7	1.41E-6	1.72E-6
1.0 (unit risk)	8.16E-6	1.65E-5	1.74E-5
100	5.58E-4	9.63E-4	1.74E-4
1,000	2.60E-2	4.22E-2	3.33E-2

[a]MLE: maximum likelihood estimate; 95% u.b.: 95% upper bound estimate. These are calculated using the alternative dose-response model.
[b]LMS: calculated by linearized multistage model (slope = 9.04 per mg/cm^2 of lung surface), using carbon core as dosimetric. Only malignant tumors are used in the calculations.

Table B-6. Excess lifetime risk for humans due to exposure to diesel exhaust emissions, under various exposure scenarios

	Alternative model[a]		
Exposure pattern	MLE	95% u.b.	LMS[b]
2.6 $\mu g/m^3$, 16 h/day, 7 days/week Normal person	1.41E-5	2.44E-5	3.00E-5
2.6 $\mu g/m^3$, 16 h/day, 7 days/week 20 pack-year smoker	2.32E-5	3.61E-5	5.38E-5
15 $\mu g/m^3$, 8 h/day, 5 days/week	3.12E-5	5.17E-5	6.18E-5

[a]MLE: maximum likelihood estimate; 95% u.b.: 95% upper bound estimate. These are calculated using the alternative dose-response model.
[b]LMS: calculated by linearized multistage model, using carbon core as dosimetric. Only malignant tumors are used in the calculations.

For the general population exposed to an ambient air concentration of 2.6 $\mu g/m^3$, the risk to normal (i.e., persons with normal respiratory functions) and smokers of 20 pack-years (as defined by Bohning et al., 1982) are provided. According to Bohning et al., the retention half-life for insoluble particle increases from 296 days for persons with normal respiratory function to 519

days for persons with a smoking history of 20 pack-years. This information is used to reduce the alveolar clearance rate in the dosimetric calculations using the same retention model that is also used to calculate dosimetric in Table B-1.

The excess lifetime risks in Tables B-5 and B-6 are calculated by actuarial life-table approach using the survival probability of the NTP control animals provided in Portier et al. (1986). Conceptually, this approach can be viewed as a weighted average of the probability of cancer occurrence over entire lifetime, weighted by survival probability. This approach is more appropriate than the one used previously in the draft report in which probability of cancer occurrence at a preselected time (730 days) is used to represent the lifetime risk; it is more appropriate because tumors here occur very late in life.

B.6. IMPLICATIONS OF THE ALTERNATIVE MODEL

Before proceeding to discuss implications of the alternative model on risk assessment, it should be noted that the parameters used in the model are estimated on the basis of high exposure concentration cancer bioassay data, not on the basis of data from laboratory measurements (e.g., mitotic rates for cells from normal and preneoplastic lesions measured over time), which usually can be obtained over lower range of exposure concentrations. Therefore, uncertainty associated with low-dose extrapolation still remains. For this reason, we will refrain from using the model to evaluate low-dose risk estimations, but rather to evaluate the relative contribution of each biological component (e.g., initiation by organic and carbon core, individually or jointly) in the model to cancer induction.

On the basis of the constructed alternative dose-response model, some specific inferences could be made from Table B-7 by changing parameter values of the original model. Table B-7 provides a comparison of risks calculated with changed parameters, assuming that animals are exposed to 7.08 mg/m^3 of diesel exhaust emissions, 7 h/day, 5 days/week (which is identical to the exposure pattern of the highest exposed group in Mauderly et al., 1987).

The following observations can be made from Table B-7.

1. When there is no diesel induced initiation (Case 2), the risk is 32% of the original model (i.e., the model without changing parameters), in contrast to 42% when exposure concentration is reduced from 7.08 to 1 mg/m^3 (not shown here). Therefore, the role of diesel-induced initiation in cancer induction increases with increasing exposure concentrations. This conclusion is intuitively obvious because spontaneous induction of initiated cells play a bigger role in cancer induction when concentration is lower. A practical implication of this

Table B-7. Effect of changing dose-dependent initiation and promotion parameters (animals are assumed to be exposed to 7.08 mg/m^3 of diesel emissions in air, 7 h/day, 5 days/week for life [i.e., the highest exposure group in Mauderly et al., 1987])

Case number	Parameters changed	Risk at 938 days	Ratio to original model
1	None (original model)	0.2067	1.00
2	a = 0 b = 0	0.0663	0.321
3	a = 0	0.1616	0.782
4	b = 0	0.1165	0.564
5	a = 2.813a[a] b = 0	0.2007	0.971
6	$\gamma_1 = 1.378\gamma_1$[a]	0.3513	1.700
7	$\gamma_1 = 1.756\gamma_1$[a]	0.4537	2.195
8	$\gamma_1 = 0$	0.0319	0.154

[a]a = 2.813a implies that a is increased by 2.813 times of its original value.

observation is that reduction of non-diesel-induced initiation (e.g., by smoking) could have greater proportion of cancer risk reduction when diesel concentration is low than when the concentration is high.

2. Cases 3 and 4 indicate that initiation by either carbon core, or organics contributes significantly to tumor incidence.

3. Case 5, along with observation Number 2 above, suggests that although diesel-induced I-cells play an important role in cancer induction, the role of initiation, however, could be assumed by either organics or carbon core alone by increasing their respective proportions. One implication is that, although existence of I-cells are important for tumor induction, these I-cells could be induced by any agent that initiates (e.g., smoking).

4. Cases 6 to 8 suggest that a small change of proliferation parameter γ could have a disproportionate change of cancer risk. Because this parameter is assumed a function of carbon core dose, lung burden overloading has a significant effect on cancer incidence. In the absence of better information, it is assumed in this report that carbon core continues to have effect at low doses.

These four observations together suggest that while effect of growth factors (which may increase value of γ) by particle overloading is important, the initiation effect of carbon core and/or organics is also essential. Although this conclusion is only tentative because the model

parameters are estimated on the basis of high concentration bioassay data, they do suggest the importance of studying the role of carbon core and organics on initiation and promotion at low- versus high-exposure concentrations. Does the relative initiation potential between organics and carbon core differ at high and low concentrations? Along with the results in Table B-6, it also suggests that a subcohort of workers who were smokers and exposed to high concentrations of diesel exhaust for a long duration would be expected to have higher lung cancer mortality.

It is interesting to observe from Table B-8 that, under the same exposure conditions, risk is greater when exposure begins later in life. This model-based conclusion is due to the fact that older animals have more spontaneously (including non-diesel) induced initiated cells that have potential to be proliferated, converted to malignant cell, and then progressing to cancer. (Note that the above observation would not contradict any observation that might show that younger animals are more sensitive to diesel exposure than older animals if a treatment induces more initiated cells in the younger animals). Assuming that the above model-generated hypothesis is realistic, an important implication is that initiation by nondiesel agents should be considered when assessing risk to humans due to exposure to diesel emissions.

B.7. CONCLUSIONS AND SUMMARY

1. The risk predictions by both alternative and LMS models are comparable over a range of exposure concentrations that is of practical interest. However, this conclusion is valid only under the assumption that the effect of carbon core on each biological component (e.g., initiation) in the model continues to exist at low doses (see further discussions about

Table B-8. Excess cancer risk to rats 200 and 300 days after termination of 6-mo exposure to D = 7.30E-5 mg/cm^2 of organics, and D = 1.89E-3 mg/cm^2 of carbon core[a]

	Days After Exposure Terminated	
Exposure Period (mo)	**200**	**300**
>6	1.48E-3	1.96E-3
6 to 12	4.01E-3	5.25E-3
12 to 18	8.32E-3	1.05E-2

[a]The lung burden is assumed to be zero over unexposed periods. It may not be a realistic assumption because the lung burden is expected to linger over the following periods after exposure terminated; however, the assumed exposure condition serves the purpose better here.

uncertainties below). Based on the Mauderly et al. (1987) study, the risks associated with continuous exposure to 1 $\mu g/m^3$ of diesel emissions calculated by two different models are summarized below:

	Alternative Model		LMS Model
Lung Tumor Data Used	**MLE**	**95% u.b.**	**95% u.b.**
Malignant tumors	8.16E-6	1.65E-5	1.74E-5
Total tumors	N.A.	N.A.	3.44E-5 (taken from Chapter 11)

2. The model suggests that populations with higher expected background rate (e.g., smokers) may be subjected to higher lung cancer risk than the populations with lower background rate. It is noted that U.S. females have about the same background lung cancer rates as the Fischer-344 rats (about 1 to 2%), whereas U.S. males have a background rate of 6%. However, because most of lung cancers are smokers, the risk to nonsmokers (males or females) should be about the same. The use of the unit risk estimate provided in Chapter 11 may somewhat underestimate risk to smokers (or other respiratory-impaired persons) unless adjustment on lung burden is made. Table B-6 provides an example of such adjustment.

B.8. DISCUSSIONS ABOUT UNCERTAINTIES OF RISK ESTIMATES

Although, it is interesting to note that risk estimates by the LMS model are comparable to those calculated by the alternative model, there are uncertainties about low-dose extrapolation by the alternative (as well as by the LMS) model: first, the model parameters are estimated statistically, not measured in the laboratory; and second, the model parameters are estimated on the basis of high-exposure data, the relationship between a parameter and exposure below the exposure range remains unknown, and the dose-parameter relationship used in the model may not be adequate for low-dose extrapolation. For instance, it is assumed that initiation rate is linearly related to doses of carbon core. Such an assumption needs be evaluated. The risk at low doses would be overestimated in this report if the relationship between initiation rate and carbon core is sublinear (concave upward). The sublinear assumption would be reasonable if there is no effect of initiation by carbon core dose (D) at low concentration. On the other hand, the risk would be underestimated if the relationship is supralinear. Therefore, it is important to evaluate how

increase of diesel-exposure concentration affects initiation rate over low-exposure concentrations. Similarly, it is important to know the relationship between dose of carbon core and cell (I-cells in particular) proliferation at low concentration.

Another aspect of uncertainty is the use of lung burdens (organics and carbon core) calculated by mathematical model, rather than actually measured. However, the impact of this uncertainty with regard to the conclusions reached in this report is not expected to be significantly altered even if the model-based dosimetrics are not accurate because the relative patterns of lung burdens between high- and low-exposure concentrations, and between animals and humans should be about the same. Although there is some observed total lung burden, these data are not used because of the following reasons.

1. The observed data are not separated by organics and carbon core.

2. There are no human data—these data are needed to predict risk in humans.

3. The observed data do not go beyond 730 days.

4. The desire is to be consistent with Chapter 11 so that results can be compared.

B.9. DISCUSSIONS ABOUT FUTURE RESEARCH NEEDS

The single most important use of a biologically based dose-response model in the cancer risk assessment is to reduce uncertainty of low-dose extrapolation when the mechanism for tumor response observed at high doses differs drastically from the low doses. However, this report can focus only on the use of the model to guide future research rather than to actually reduce uncertainty of risk estimate because of our inability to obtain biological parameters in the model. If a chemical is known to induce disproportionately larger cell proliferation (in normal, initiated, and/or malignant cells) at high doses than at low doses, then a model that reflects this fact would be useful. With this in mind, our effort should be to identify "components" of carcinogenesis (e.g., increase of mitotic rate) that are disproportionately more affected at high doses than at low doses and to develop models that incorporate those high-dose effects. For the diesel risk assessment, the "components" that require further study include effects of organics and carbon core, individually or jointly, on initiation, proliferation, conversion, and progression steps of carcinogenesis. In order to use biologically based models of carcinogenesis in risk assessments, one needs to know the relationship between parameter values in a model and exposure (or dose). Ideally, some of these parameters, if not all, should be measured directly in the laboratory, or indirectly estimating from neoplastic and preneoplastic lesions (e.g., number of foci, adenomas, and tumors in a lung).

Cell proliferation is an increase in the cell population of different stages: normal, initiated, or malignant cells. Enhanced cell proliferation of normal target cells may itself increase the frequency of mutations, either by inducing error in replication or by converting DNA adducts to mutations before DNA repair can occur. The model implies that tumor incidence is linearly proportional to initiation rate. On the other hand, enhanced cell proliferation of initiated cells could lead to more than linear increase of tumor incidence. Therefore, proliferation of I-cells has a greater impact on tumor incidence than proliferation of normal cells. However, this does not mean that initiation potential of compounds (organics or carbon core) is not important. As discussed previously, it is important to determine the ability of these compounds to initiate at low versus high doses; this has a significant implication for low-dose extrapolation. From the viewpoint of mathematical modeling, cell proliferation is the result of a decrease of cell death rate and/or an increase of mitotic rate, regardless of underlying biological mechanisms. Therefore, it is logical to construct a model (as is done here) with a proliferation component involving cell death and mitosis, and important to obtain data at the cellular level even if biological mechanism at the molecular level is not yet known. If a more precise mechanism is known and the quantitative data are available, then the proliferation component of the model can be improved by incorporating the available biological information. Most of the two-stage models consider a single malignant cell to be equivalent to a tumor. If a compound is known to affect the cell proliferation of tumor cell population, a model that incorporates tumor progression should be used. For the diesel modeling, we assume that particles could enhance the proliferation of malignant cells. This assumption needs to be verified. Another model-generated hypothesis is that persons with higher number of initiated cells are subjected to higher lung cancer risk when exposed to diesel emissions. (A person could have a higher number of initiated cells due to exposure to diesel and/or nondiesel agents, or simply by acquiring more spontaneously induced initiated cells through aging).

In summary, information that is necessary to construct a biologically based dose-response model includes (1) identifying roles that are played by organics and carbon core (individually or jointly) with respect to initiation, proliferation, conversion, and progression, at low versus high doses; (2) quantitative measurements of cellular dynamics (e.g., mitotic rate) for cells at different stages and exposure concentrations; and (3) relationship between parameters and exposure or dose. Because many biological parameters are expected to be age-dependent, they should be measured over different time points. Furthermore, frequency and size of preneoplastic foci, nodules, and tumors could also provide useful information toward improving risk assessment. Some of these data may be obtained by initiation-promotion type of study.

B.10. RFERENCES

Bohning D., Atkins, H., and Cohn, S. Long-term particle clearance in man: normal and impaired. Ann. Occup. Hyg. 26:259-271.

Chen, C., and Farland, W. Incorporating Cell Proliferation in Quantitative Cancer Risk Assessment: Approach, Issues, and Uncertainties. In: Chemically Induced Cell Proliferation: Implications for Risk Assessment, B. Butterworth, T. Slaga, W. Farland, and M. McClain, eds.; pp. 481-499 (Wiley-Liss, Inc., New York, 1991).

Cohen S. and Ellwein L. Cell growth dynamics in long-term bladder carcinogenesis. Toxicol. Letters, 43:151-173, 1988.

Dempster, A., Laird, N., and Rubin, D. Maximum likelihood from incomplete data via the EM algorithm. Roy. Statist. Soc., B 39:1-38, 1977.

Greenfield, R., Ellwein, L., and Cohen, S. A general probabilistic model of carcinogenic analysis experimental bladder cancer, Carcinogenesis 5:437-445, 1984.

Kauffman S. Kinetics of alveolar epithelial hyperplasia in lungs of mice exposed to urethane I. quantitative analysis of cell populations, Laboratory Investigations, 30:170-175, 1974.

S. Moolgavkar and D. Venzon, "Two-Event Models for Carcinogenesis: Incidence Curve for Childhood and Adult Tumors," Math. Biosciences 47, 55-77 (1979).

S. Moolgavkar and A. Knudson, "Mutation and Cancer: A Model for Human Carcinogenesis," Journal of the National Cancer Institute 66, 1037-1052 (1981).

Portier C, Hedges, J., and Hoel, D. Age-specific models of mortality and tumor onset for historical control animals in the National Toxicology Program's carcinogenicity experiments. Cancer Research 46:4372-4378, 1986.

W. Tan and C. Chen, A nonhomogeneous stochastic model of carcinogenesis for assessing risk of environmental agents. To appear in: Mathematical Population Dynamics, Proceedings of the Third International Conference, Pau, France, 1992.

APPENDIX B-1
E-M ALGORITHM

The E-M algorithm is derived below. It will be used to calculate maximum likelihood estimate of parameters of the alternative model. Data used for the E-M algorithm is taken from Mauderly et al. (1987), which includes time when an animal died (natural mortality or sacrifice) with or without (malignant) tumors. The computer program for the E-M calculations was developed by Mr. Daliang Chang of the Computer Science Corporation under an EPA contract. The theory of E-M algorithm can be found in Dempster et al. (1977).

Assume that the distinct times when animals died by either natural mortality or sacrifice are $t_1 < t_2 < ... < t_m$. The observations can be classified as follows:

$a_{1x}(I)$: observed number of natural deaths without tumor at time t_i in the treatment group x (There are four groups for diesel data [i.e., x = 1, 2, 3, 4.]),

$a_{2x}(I)$: observed number of natural deaths with tumor at time t_i in the treatment group x,

$b_{1x}(I)$: series sacrifice at time t_i without tumor in the treatment group x,

$b_{2x}(I)$: series sacrifice at time t_i with tumors in the treatment group x.

Let T_d represent the time an animal died and T the time a tumor developed.

$\alpha_x(I) = Pr\{T_d = t_i | T_d \geq t_i, T > t_i, x\}$ (conditional probability of death without tumor)

$\beta_x(I|u) = Pr\{T_d = t_i | T_d \geq t_i, T \in (t_{u-1}, t_u], x\}$ (related to deaths with tumors)

Define,

$$A_x(i) = \prod_{j=1}^{i} [1 - \alpha_x(j)],$$

$$B_x(i|u) = \prod_{j=u}^{i} [1 - \beta_x(j|u)].$$

$$S_x(t) = Pr\{T \geq t | x\} = \exp[-\int_{t_0}^{t} h(x)dx].$$

The function $S_x(t)$ is the probability of tumor free by time t. The exact form of the hazard function h(x) and $S_x(t)$ are given in the next section.

Let

$a_{2x}(I|u)$ = number of natural death at t_i with tumor developed during $(t_{u-1}, t_u]$, in the treatment group x, u < I,

$b_{2x}(I|u)$ = number of sacrifice at t_i with tumor developed during $(t_{u-1}, t_u]$, in the treatment group x, u < I,

Then

$$a_{2x}(i) = \sum_{u=1}^{i} a_{2x}(i|u), \; b_{2x}(i) = \sum_{u=1}^{i} b_{2x}(i|u)$$

Let

$$P_x(i|u) = \frac{A_x(u)S_x(t_u)\Delta_x(t_u)B_x(i-1|u)\beta_x(i|u)}{\sum_{j=1}^{i} A_x(j)S_x(t_j)\Delta_x(t_j)B_x(i-1|j)\beta_x(i|j)},$$

and

$$Q_x(i|u) = \frac{A_x(u)S_x(t_u)\Delta_x(t_u)B_x(i|u)}{\sum_{j=1}^{i} A_x(j)S_x(t_j)\Delta_x(t_j)B_x(i|j)},$$

where

$$\Delta_x(t_i) = \frac{S_x(t_{i-1})}{S_x(t_i)} - 1.$$

Given $a_{2x}(x)$, $\{a_{2x}(i|u), u = 1, ..., i\}$, is an $(i - 1)$-dimension multinomial with parameter $\{a_{2x}(i), P_x(i|u), u = 1, ..., i\}$.

Thus, $$E[a_{2x}(i|u)|a_{2x}(i)] = a_{2x}(i)P_x(i|u).$$

Similarly, $\{b_{2x}(i|u), u = 1, ..., i\}$, is an $(i - 1)$-dimension multinormial with parameters $\{b_{2x}(i), Q_x(i|u), u = 1, ..., i\}$, and

$$E[b_{2x}(i|u)|b_{2x}(i)] = b_{2x}(i)Q_x(i|u).$$

It can be shown that the likelihood function is proportional to

$$L = \prod_x \prod_{i=1}^{m} [S_x(t_i)]^{a_{1x}(i)+b_{1x}(i)+m_x(i)}[\Delta_x(t_i)]^{m_x(i)},$$

where

$$m_x(i) = \sum_{u=i}^{m} [a_{2x}(u|i) + b_{2x}(u|i)] .$$

Let

$$R_{1x}(i) = \sum_{j=i}^{m} [a_{1x}(j) + b_{1x}(j) + m_x(j)], \text{ and}$$
$$R_{2x}(i|u) = \sum_{j=i}^{m} [a_{2x}(j|u) + b_{2x}(j|u)] .$$

Let

$$\Theta = (\mu_1, \mu_2, \gamma_1, ...)$$

be a vector of parameters in function S;

$$\alpha_x = [\alpha_x(1),\ \alpha_x(2),\ ...,\ \alpha_x(m)],\ \text{and}$$
$$\beta_x(u) = [\beta_x(1|u),\ \beta_x(2|u),\ ...,\ \beta_x(m|u)]$$

be vectors of parameters related to conditional probabilities of death with and without tumors. These parameters, along with those in Θ_x will be estimated by the E-M algorithm described below.

The M-step:

Given initial values $a_{2x}(i|u)$ and $b_{2x}(i|u)$, estimate

1. $\overline{\alpha_x}(i) = a_{1x}(i)/R_{1x}(i)$ and
2. $\overline{\beta_x}(i|u) = a_{2x}(i|u)/R_{2x}(i|u)$, and
3. obtain $\overline{\theta_x}$ by maximizing the log of L.

The E-Step:

Given the estimated values on $\alpha_x(i)$, $\beta_x(i)$, and Θ_x from the M-step, compute $P_x(i|u)$ and $Q_x(i|u)$, and obtain estimates of $a_{2x}(i|u)$ and $b_{2x}(i|u)$ by

$$\overline{a_{2x}}(i|u) = a_{2x}(i)\overline{P_x}(i|u),\ \text{and}$$
$$\overline{b_{2x}}(i|u) = b_{2x}(i)\overline{Q_x}(i|u).$$

With the estimated values of $a_{2x}(i|u)$ and $b_{2x}(i|u)$ available from the E-step, go back to the M-step. Repeat the same process until estimates are stabilized.

APPENDIX B-2
A TUMOR GROWTH MODEL

The tumor growth model with piece-wise constant parameters is taken from Tan and Chen (1992), which is an extension of a stochastic model developed by Chen and Farland (1991). This model has a similar biological motivation as the two-stage model proposed by Greenfield et al. (1984), which has been used by Cohen and Ellwein (1988) to analyze bladder tumors. However, the two models differ from each other with respect to their mathematical formulations; the one adopted in this report is a stochastic model, whereas the other is a deterministic model and does not allow for parameters estimation because the model does not have complete mathematical expression.

Although its most general form will not be used here because of the lack of data, it is worthwhile to note that the stochastic model by Chen and Farland (1991) has two desirable features: (1) it allows for any cell growth distributions (e.g., Gompertz), rather than limiting only to the exponential distribution as in other existing models; and (2) it incorporates the birth and death of tumor cells, rather than assuming that a tumor is born once a single tumor cell occurs, an assumption made by the MVK model (a model proposed by Moolgavkar et al., 1979, 1981). Therefore, if information on cell lifetime distribution, and the progression phase of tumor development is available, a reasonably realistic model can be constructed.

For completeness of the report, a brief description of the model will be presented here. The following notations are needed for the model:

N(t): number of normal (target) cells at time t,

μ_1: initiation rate, and

f(t): the probability density function for the lifetime of an initiated cell (I-cell).

For an I-cell, at the end of its lifetime it either divides (mitosis) or dies (programmed or nonprogrammed death). If it enters into mitosis, it either divides into two I-cells with probability α, or divides into one I-cell and one malignant cell (M-cell) with probability μ_2. Note that, at the end of a cell's lifetime, the probability for the cell to die is $\beta = 1 - \alpha - \mu_2$. A similar setup (i.e., to allow for any cell lifetime distribution) can be made for an M-cell. However, we will confine ourselves to a simpler version assuming that an M-cell lifetime follows an exponential distribution.

Thus, we can simply assume that an M-cell follows a simple birth-death process; it can either divide into two M-cells with a rate α_m or die with a rate β_m.

When parameters are constant over time (ages), the hazard function is given by

$$h(t) = \mu_1\mu_2\int_0^t N(s)m(t - s)\, ds$$

where

$$m(t) = \frac{(y_2 - y_1)^2\exp[A(t)\alpha(y_2 - y_1)]}{\langle(1 - y_1) + (y_2 - 1)\exp[A(t)\alpha(y_2 - y_1)]\rangle^2};$$

where $y_1 < y_2$ are two real roots of $\alpha y^2 - (\alpha + \beta + \mu_2 q)y + \beta = 0$; $\alpha + \beta + \mu_2 = 1$; $q = 1 - \beta_m/\alpha_m$; $A(t) = \int a(x)dx$, where $a(t) = f(t)/[1 - F(t)]$ is the hazard function of the cell lifetime, F(t) is the cumulative function of f(t). Two special cases of interest here are $a(t) = \gamma$, when the exponential distribution is assumed; and $a(t) = \exp(-\gamma t)$, when the Gompertz distribution is assumed.

When exponential distribution (i.e., $a(t) = \gamma$, or $A(t) = \gamma t$) and $q = 1$ are assumed, the model is equivalent to the MVK model. A special case that may be more appropriate than the exponential distribution is when the Gompertz distribution is assumed (i.e., $A(t) = [1 - \exp(-\gamma t)]/\gamma$).

For the model with time-dependent parameters, assume that the study begins at time t_0. Divide time scale $(t_0, t]$ into k subintervals $L_j = (t_{j-1}, t_j]$, $j = 1, 2, ...k-1$ and $L_k = (t_{k-1}, t_k]$ where $t_k = t$ (note that these subintervals may not be the same subintervals defined by deaths or sacrifice before). The parameters that vary over subintervals $(t_{i-1}, t_i]$, $i = 1, 2, ..., k$ are μ_{1j}, α_j, β_j, μ_{2j}, N_j, and those parameters related to f(t). The hazard function is given by

$$h(t) = \sum_{j=1}^{k}[\mu_{1j}\mu_{2j}N_j\int_{t_{j-1}}^{t_j} m_j(t_j - s)ds]\prod_{i=j+1}^{k} m_j(t_i - t_{i-1}),$$

where

$$\prod_{i=j+1}^{k} m_j(t_j - t_{j-1}) = 1, \text{ when } j = k$$

and

$$m_j(t) = \frac{(y_{2j} - y_{1j})^2 \exp[A_j(t)\alpha_j(y_{2j} - y_{1j})]}{\langle (1 - y_{1j}) + (y_{2j} - 1)\exp[A_j(t)\alpha_j(y_{2j} - y_{1j})]\rangle^2},$$

where $y_{1j} < y_{2j}$ are two real roots of $\alpha_j y^2 - (\alpha_j + \beta_j + \mu_{2j}q_j)y + \beta_j = 0$; $\alpha_j + \beta_j + \mu_{2j} = 1$; $q_j = 1 - \beta_{mj}/\alpha_{mj}$, $j = 1, 2, ..., k$.

When exponential distribution (i.e., $A_j(t) = \gamma_j t$ and $q_j = 1$ are assumed, the model is equivalent to the MVK model with piece-wise constant parameters. A special case that may be more appropriate than the exponential distribution is when the Gompertz distribution is assumed (i.e., when $A_j(t) = \{1 - \exp[-\gamma_j t]\}/\gamma_j$).

For the diesel alternative model, the total time is divided into five (i.e., $k = 5$) subintervals. It is shown in Tan and Chen (1992) that, under the assumption of exponential cell lifetime distribution, the tumor free distribution function, $S_x(t)$, can be written as

$$S(t) = \exp\{-\sum_{j=1}^{k}[A_{jj}(t_{j-1}, s_j) + \sum_{i=1}^{j-1} A_{ij}(t_{i-1}, s_j)]\}$$

where $s_j = t_j$ if $j < k$ and $s_j = t$ if $j = k$, and

$$A_{jj}(t_{j-1}, s_j) = 2N_j\mu_{1j}\mu_{2j}\frac{1}{w_I + z_I}\{-(s_j - t_{j-1}) + \frac{2}{\gamma_{Ij}(w_I - z_I)}\log[1 + \frac{w_I - z_I}{2w_I}(e^{w_I\gamma_{Ij}(s_j - t_{j-1})} - 1)]\},$$

$$A_{ij}(t_{i-1},s_j) = 4N_i\mu_{1i}\mu_{2j}[\frac{1}{\gamma_{Ii}(w_I^2 - z_I^2)}]x \langle\log[\frac{w_I + z_I + (w_I - z_I)\exp(w_I\Delta_{i+1,j-1}(t_i,t_{j-1})}{w_I + z_I + (w_I - z_I)\exp(w_I\Delta_{i,j-1}(t_{i-1},t_{j-1})}] - \log[\frac{w_I + z_I + (w_I - z_I)\exp(w_I\Delta_{i+1,j}(t_i,s_j)}{w_I + z_I + (w_I - z_I)\exp(w_I\Delta_{ij}(t_{i-1},s_j)}]\rangle,$$

where,

$W_I = [(\alpha + \beta + \mu_2 q)^2 - 4\alpha\beta]^{1/2}$,

$Z_I = \alpha - \beta - \mu_2 q$, and

$\triangle_{ii}(s,t) = \gamma_i(t - s)$ if both s and t are in the same closed subinterval $[t_{i-1}, t_i]$ and

$$\Delta_{ij}(s,t) = \gamma_i(t_i - s) + \sum_{r=i+1}^{j-1} \gamma_r(t_r - t_{r-1})$$

if $s \epsilon L_i$, $t \epsilon L_j$ with $t_i < t_j$